

In the Claims:

Please cancel claims 1-12, without prejudice.

Please add the following new claims:

13. A text-to-speech conversion system for interlocking with multimedia comprising:

- a multimedia information input unit for organizing text, prosody information, information on synchronization with a moving picture, lip-shape information, picture information, and individual property information;
- a data distributor by each media for distributing the information of said multimedia information input unit into information for each media;
- a language processor for converting the text distributed by said data distributor by each media into a phoneme stream, presuming prosody information and symbolizing the presumed prosody information;
- a prosody processor for calculating a prosody control parameter value from the symbolized prosody information;
- a synchronization adjustor for adjusting a duration of each phoneme using the synchronization information distributed by said data distributor by each media;
- a synthesis unit database for receiving the individual property information from said data distributor by each media, selecting synthesis units adaptable to gender and age, and outputting data required for synthesis;

a signal processor for producing a synthesized speech using the prosody control parameter and the data output from said synthesis unit database; and

a picture output apparatus for outputting the picture information distributed by said data distributor by each media on to a screen.

14. A method for organizing input data of a text-to-speech conversion system for interlocking with multimedia, said method comprising the steps of:

(a) classifying multimedia input information organized for enhancing natural synthesized speech and implementing synchronization of multimedia with text-to-speech into text, prosody information, information on synchronization with a moving picture, lip-shaped information, picture information, and individual property information using a multimedia information input unit;

(b) distributing using a data distributor by each media the multimedia input information classified in the multimedia information input unit based on respective information;

(c) converting the text distributed by the data distributor by each media into a phoneme stream, presuming prosody information and symbolizing the presumed prosody information using a language processor;

(d) calculating a prosody control parameter value other than a prosody control parameter included in the multimedia input information using a prosody processor;

(e) adjusting a duration of each phoneme using a synchronization adjustor so as to synchronize a processing result of the prosody processor with a picture signal according to the synchronization information distributed by the data distributor by each media;

(f) selecting synthesis units adaptable to gender and age based on the individual property information from the data distributor by each media using a synthesis unit database and outputting data required for synthesis;

(g) producing synthesized speech using a signal processor based on the prosody information distributed by the data distributor by each media, a processing result of the synchronization adjustor, and the data from the synthesis unit database; and

(h) outputting the picture information distributed by the data distributor by each media onto a screen using a picture output unit.

3 ~~14~~
15. The method in accordance with claim ~~14~~, wherein the organized multimedia information comprises text information, prosody information, information on synchronization with a moving picture, lip-shaped information, and individual property information.

24 ~~15~~
16. The method in accordance with claim ~~15~~, wherein the prosody information comprises a number of phoneme, phoneme stream information, duration of each phoneme, pitch pattern of the phoneme, and energy pattern of the phoneme.

~~17~~⁵. The method in accordance with claim ~~16~~², wherein the duration time of the phoneme is indicative of a value of pitch at a beginning point, a mid point, and an end point within the phoneme.

~~18~~⁶. The method in accordance with claim ~~17~~⁵, wherein the energy pattern of the phoneme is indicative of a value of energy in decibels at the beginning point, the mid point, and the end point within the phoneme.

~~19~~¹. The method in accordance with claim ~~15~~³, wherein the synchronization information comprises text, lip-shape, location information with a moving picture, and duration information.

~~20~~⁸. The method in accordance with claim ~~16~~³, wherein the synchronization information comprises a beginning point, duration and delay time information of a starting point, and duration of each phoneme is controlled by the synchronization information.

~~21~~⁹. The method in accordance with claim ~~15~~³, wherein the synchronization information is composed of a duration of a beginning point of a sentence, a duration information of a starting point, and duration of each phoneme is controlled by forecast lip-shape considered an articulation manner of the phoneme and articulation control of lip-shape within the synchronization and duration information of the synchronization information.

22.¹⁰ The method in accordance with claim ³15, wherein the synthesized speech is produced based on beginning point information, end point information, and phoneme information for each phoneme within an interval associated with a speech signal.

23.¹¹ The method in accordance with claim ³15, wherein the synthesized speech is produced based on a distance of an opening between an upper lip and a lower lip, a distance between end points of the lips, and an extent of projection of a lip, and a lip-shape quantized and normalized pattern is defined depending on articulation location and articulation manner of the phoneme on a basis of pattern with discriminative property.

24.¹² The method in accordance with claim ³15, wherein if the multimedia input information comprises prosody information, further comprising the steps of:

(i) converting the prosody information into a data structure recognizable by the signal processor; and

(j) transmitting the converted prosody information the prosody processor and the synchronization adjustor.

25.¹³ The method in accordance with claim ³15, wherein if the multimedia input information includes individual property information, further comprising the steps of:

(k) converting the individual property information into a data structure recognizable by the synthesis unit database and the prosody processor within the text-to-speech;